Glaucoma Detection based on Deep Convolutional Neural Network

Xiangyu Chen[1], Yanwu Xu[1], Damon Wing Kee Wong[1],
Tien Yin Wong[2], and Jiang Liu[1]

***Abstract*— Glaucoma is a chronic and irreversible eye disease, which leads to deterioration in vision and quality of life. In this paper, we develop a deep learning (DL) architecture with convolutional neural network for automated glaucoma diagnosis. Deep learning systems, such as convolutional neural networks (CNNs), can infer a hierarchical representation of images to discriminate between glaucoma and non-glaucoma patterns for diagnostic decisions. The proposed DL architecture contains six learned layers: four convolutional layers and two fully-connected layers. Dropout and data augmentation strategies are adopted to further boost the performance of glaucoma diagnosis. Extensive experiments are performed on the *ORIGA* and *SCES* datasets. The results show area under curve (AUC) of the receiver operating characteristic curve in glaucoma detection at 0.831 and 0.887 in the two databases, much better than state-of-the-art algorithms. The method could be used for glaucoma detection.**

I. INTRODUCTION

Glaucoma is one of the common causes of blindness, and is predicted to affect around 80 million people by 2020 [16]. It is a chronic eye disease that leads to vision loss, in which the optic nerve is progressively damaged. As the symptoms only occur when the disease is quite advanced, glaucoma is called the silent thief of sight. Although glaucoma cannot be cured, its progression can be slowed down by treatment. Early detection of glaucoma based on effective images is highly needed.

Digital Fundus Image is one of the main and popular modalities to diagnose glaucoma. Since it is possible to acquire DFIs in a noninvasive manner which is suitable for large scale screening, DFI has emerged as a preferred modality for large-scale glaucoma screening. In a glaucoma screening program, an automated system decides whether or not any signs of suspicious for glaucoma are present in an image. Only those images deemed suspect by the system will be passed to ophthalmologists for further examination.

Glaucoma is detected basically by utilizing the medical history, intra-ocular pressure and visual field loss tests together with a manual assessment of the Optic Disc (OD) through ophthalmoscopy. OD is the location where ganglion cell axons exit the eye to form the optic nerve, through which visual information of the photo-receptors is transmitted to the brain. In 2D images, the OD can be divided into two distinct zones: a central bright zone called the optic cup (in short, cup) and a peripheral region called the neuroretinal rim. The loss in optic nerve fibres leads to a change in the structural appearance of the OD, namely, the enlargement of cup region (thinning of neuroretinal rim) called cupping. Since one of the important indicators is the enlargement of the cup with respect to OD, various parameters are considered and estimated to detect the glaucoma, such as the vertical cup to disc ratio (CDR) [17], disc diameter [5], ISNT rule [6], and peripapillary atrophy (PPA) [14].

Among the structural image cues studied for glaucoma diagnosis, CDR is a major consideration of clinicians [2][12][13]. However, clinical assessment by manually annotating the cup and disc for each image is labor-intensive, and automatically segmenting the disc and cup in fundus images is also time consuming. Extracting the optic disc region of interest (ROI) will produce a smaller initial image which takes much lesser time taken to process compared to segmenting disc and cup [3]. In this paper, we consider the ROI image as the input of the proposed deep Convolutional Neural Network (CNN).

For glaucoma detection, the disease pattern in DFIs is complex and hidden, which is different from the natural scene images. The analysis task of natural scene images are related to object detection of regions that has an obvious visual appearance (e.g. texture, shape or color). But glaucoma disease patterns could be only observed by the training and expertise of the examiner.

Deep learning (DL) is an active research topic which learns discriminative representations of data. The DL architectures are formed by the composition of multiple linear and non-linear transformations of the data, with the goal of yielding more abstract and ultimately more useful representations [15]. Convolutional neural networks (CNNs) are deep learning architectures, are recently been employed successfully for image segmentation and classification tasks [18][15][9]. DL architectures are an evolution of multilayer neural networks (NN), involving different design and training strategies to make them competitive. These strategies include spatial invariance, hierarchical feature learning and scalability [15]. In this paper, effectively capturing the deep features of glaucoma based on deep CNN is our main interest.

Therefore, we are motivated to propose a deep learning framework for capturing the discriminative features that better characterize the hidden patterns related to glaucoma. The adopted DL structure consists of six layers:four convolutional layers and two fully-connected layers, which infers a hierarchical representation of images to discriminate between glaucoma and non-glaucoma patterns for diagnostic

[1]Xiangyu Chen, Yanwu Xu, Damon Wing Kee Wong and Jiang Liu are with the Institute for Infocomm Research, Agency for Science, Technology and Research, 138632, Singapore {chenxy, yaxu, wkwong, jliu} at i2r.a-star.edu.sg

[2]Tien Yin Wong is with the Singapore National Eye Centre, 168751, Singapore tien_yin_wong at nuhs.edu.sg

978-1-4244-9270-1/15/$31.00 ©2015 IEEE

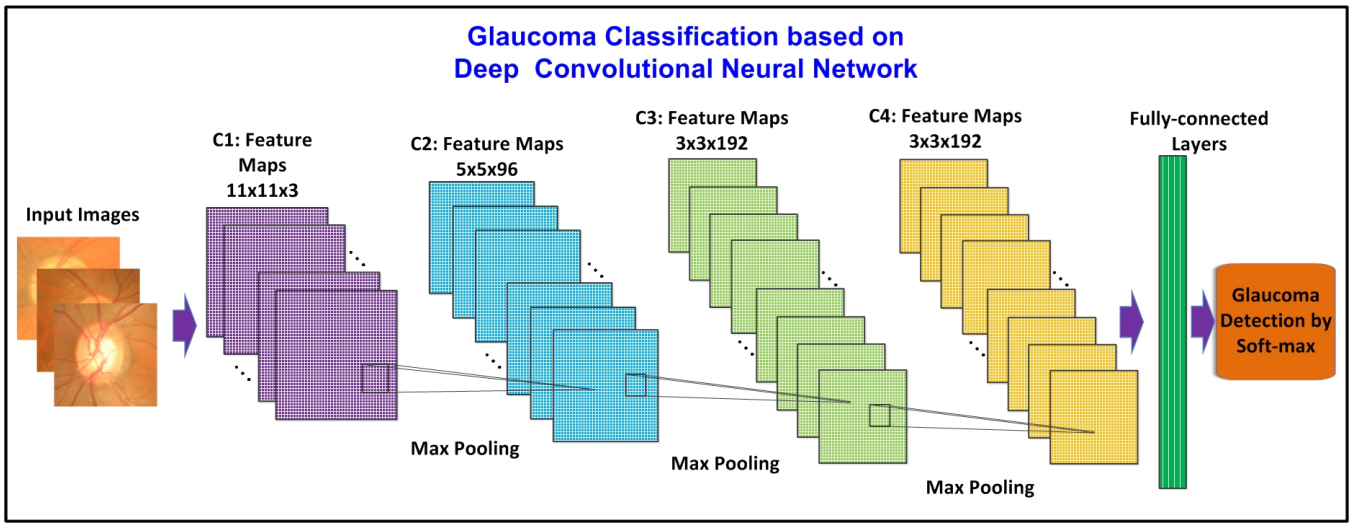

Fig. 1. System overview of our proposed glaucoma detection based on deep convolutional neural network. The proposed DL architecture contains six learned layers: four convolutional layers and two fully-connected layers. The output of the last two fully-connected layer is fed to a soft-max classifier for glaucoma prediction. Response-normalization layers follow the first and second convolutional layers. Max-pooling follow both response-normalization layers as well as convolutional layers. The dimension of input image is $256 \times 256 \times 3$. Data Augmentation is done by extracting random 224×224 patches from the input images. For better viewing, please see the color pdf file.

decisions. In addition, to reduce the overfitting problem, we adopt response-normalization layers and overlapping-pooling layers. In order to further boost the performance, dropout and data augmentation strategies are also adopted in the proposed DL architecture. The paper is organized as follows. In Section I, we have given an introduction of the background and motivation for the method. In Section II, we introduce the overview of the deep learning architecture. In Section III, we introduce the glaucoma classification based on CNN. Section IV shows the experimental results, followed by the conclusions in the last section.

II. Overview of the Deep Learning Architecture

In this paper, the proposed deep learning architecture is based on CNN. As shown in Fig.1, the net of CNN contains six layers with weights: the first four are convolutional and the remaining two are fully connected. The output of the last fully-connected layer is fed to a soft-max classifier for glaucoma prediction. Response-normalization layers and overlapping layers are employed in our proposed learning architecture as in [18].

A. Convolutional Layers

Convolutional layers are usually employed to learn small feature detectors based on patches randomly sampled from an image. A feature in the image at some location can be calculated by convolving the feature detector and the image at that location.

Let us denote $\mathbf{H}^{(n-1)}$ and $\mathbf{H}^{(n)}$ as the input and output for the n-th layer of CNN. Let $\mathbf{H}^{(0)}$ be the 2D input image patch and $\mathbf{H}^{(N)}$ be the output of the last layer N. Let $M_I^{(n)} \times M_I^{(n)}$ and $M_O^{(n)} \times M_O^{(n)}$ be the size of the input and the output map, respectively, for layer n. Denote $Q_I^{(n)}$ and $Q_O^{(n)}$ as the number of input and output maps respectively for the n-th layer. Since the input to n-th layer is the output of $(n-1)$-th layer, $Q_I^{(n)} = Q_O^{(n-1)}$ and $M_I^{(n)} = M_O^{(n-1)}$. Let $\mathbf{H}_j^{(n)}$ be the j-th output-feature-map of n-th lay. As the input-feature-maps of the n-th layer are actually the output-feature-maps of the $(n-1)$-th layer, $\mathbf{H}_i^{(n-1)}$ is the i-th input-feature-map of layer n. Finally, we could get the output of a convolutional layer:

$$\mathbf{H}_j^{(n)} = f(\Sigma_i \mathbf{H}_i^{(n-1)} * \mathbf{w}_{ij}^{(n)} + b_j^{(n)} \mathbf{1}_{M_O^{(n)}}), \tag{1}$$

where $\mathbf{w}_{ij}^{(n)}$ is the kernel linking i-th input map to j-th output map, $*$ denotes the convolution, $0 \le i \le Q_I^{(n-1)}, 0 \le j \le Q_O^{(n-1)}$, and $b_j^{(n)}$ is the bias element for j-th output-feature-map of n-th layer.

B. Response-normalization Layers

Response-normalization layers follow the first and second convolutional layers in the proposed deep learning architecture. In a neural network, the traditional way to compute a neuron's output$f$ utilizing $f(x) = tanh(x)$ (x is the input). However, these saturating nonlinearities are much slower than the non-saturating nonlinearity $f(x) = \max(0,x)$. The neurons with this nonlinearity are called Rectified Linear Units (ReLUs). Denoting the activity of a neuron computed by applying kernel i at position (x,y) as $u_{x,y}^i$, and after performing the ReLU nonlinearity on it, the response-normalized activity is computed as:

$$\vec{u}_{x,y}^i = u_{x,y}^i / (\alpha + \lambda \sum_{j=\max(0,i-p/2)}^{\min(Z-1,i+p/2)} (u_{x,y}^j)^2)^\eta, \tag{2}$$

where Z is the total number of kernels in the layer, the sum $\sum_{j=\max(0,i-p/2)}^{\min(Z-1,i+p/2)} (u_{x,y}^j)^2$ computes p adjacent kernel maps at the same spatial position (x,y). The constants α, p, λ and

η are hyper-parameters whose values are determined using a validation set.

C. Overlapping-pooling Layers

Pooling layers in CNNs summarizes the statistics of a feature over a region in the image. A pooling layer consists of a grid of pooling units spaced p pixels apart, each summarizing a neighborhood of size $s \times s$ centered at the location of the pooling unit. When $p = s$, we obtain traditional local pooling. When $p < s$, we obtain overlapping pooling,which could help to overcome the overfilling. In our learning architecture, max-pooling layers follow both response-normalization layers as well as the convolutional layer as shown in Fig.1.

III. Glaucoma Classification based on CNN

As shown in Fig.1, the output of the last two fully-connected layer is fed to a soft-max classifier for glaucoma prediction.

A. Region of Interest (ROI) Extraction

In this paper, we use the ROI image as the input of the proposed deep Convolutional Neural Network (CNN), which produces a smaller initial image which takes much lesser time taken to process compared to segmenting disc and cup [3]. In the previous approach in ARGALI [8], the retinal fundus image is divided into grids and the region of interest (ROI) where the optic nerve head is located is determined by finding the grid which is the mostly likely candidate. Since correctly identifying the ROI speed up the calculation of the process and boost up the performance of glaucoma detection, we adopt the algorithm in [3] to accurately locate the ROI in the retinal fundus image.

In the method of [3], a preprocessing step was introduced to reduce or remove the bright fringe, which involves finding the center of the trimming circle and the trim radius. After obtaining ROI, the extracted result is down-sampled to a fixed resolution of 256×256. Finally, the mean value over all the pixels in the disc image is subtracted from each pixel to remove the influence of illumination variation among images.

B. Dropout and Data Augmentation

In order to overcome the overfitting on image data, we employ data augmentation to artificially enlarge the dataset using label-preserving transformations, and dropout for model combination .

Dropout: We use dropout in the two fully-connected layers in our proposed deep learning architecture. Dropout consists of setting to zero the output of each hidden neuron with probability 0.5 [10]. If the neurons in CNN are dropped out, they do not contribute to the forward pass and do not participate in back propagation. During testing, we use all the neurons but multiply their outputs by 0.5.

Data Augmentation: If we do not adopt data augmentation, our network will suffer from substantial overfitting. Data augmentation consists of generating image translations and horizontal reflections[11]. At training time, data augmentation is performed by extracting random 224×224 patches including their horizontal reflections from the 256×256 images, and training our network on these extracted patches. At test time, the CNN makes a prediction by extracting five 224×224 patches including the four corner patches and the center patch, as well as their horizontal reflections, and averaging the predictions made by the network soft-max layer on these ten patches. Let us denote this test strategy as multi-view test (MVT).

IV. Experiments

To evaluate the glaucoma diagnosis performance of our proposed deep CNN method, we perform experiments on two glaucoma fundus image datasets ORIGA [1] and SCES [4].

A. Evaluation Criteria

In this work, we utilize the area under the curve (AUC) of receiver operation characteristic curve (ROC) to evaluate the performance of glaucoma diagnosis. The ROC is plotted as a curve which shows the tradeoff between sensitivity TPR (true positive rate) and specificity TNR (true negative rate), defined as

$$TPR = \frac{TP}{TP+FN}, \; TNR = \frac{TN}{TN+FP}, \tag{3}$$

where TP and TN are the number of true positives and true negatives, respectively, and FP and FN are the number of false positives and false negatives, respectively.

B. Experimental Setup

We adopt the same settings of the experiments for glaucoma diagnosis in [7] in this work to facilitate comparisons. The ORIGA dataset with clinical glaucoma diagnoses, is comprised of 168 glaucoma and 482 normal fundus images. The SCES dataset contains 1676 fundus images, and 46 images are glaucoma cases.

C. Experimental Results

In order to validate the effectiveness of our deep CNN on glaucoma diagnosis accuracy, we compare the predictions of CNN to state-of-the-art reconstruction-based method [7]. For *ORIGA* dataset, we adopt the same setting of [7]. The training set contains a random selection of 99 images from the whole 650 images, and the remaining 551 images are used for testing. For *SCES* dataset, we use the 650 images from *ORIGA* for training, and the whole 1676 images of *SCES* are the test data. The AUC values of our method on *ORIGA* and *SCES* are 0.831 and 0.887, respectively. For the state-of-the-art reconstruction-based method, the AUC values are 0.823 and 0.860, respectively

In addition, Fig. 2 gives six sample results by our proposed algorithm. Each fundus image is diagnosed by clinicians and the predicted labels with probabilities by our algorithm. Though the third row has low image quality, our method still detects the glaucoma, indicating its robustness and stability.

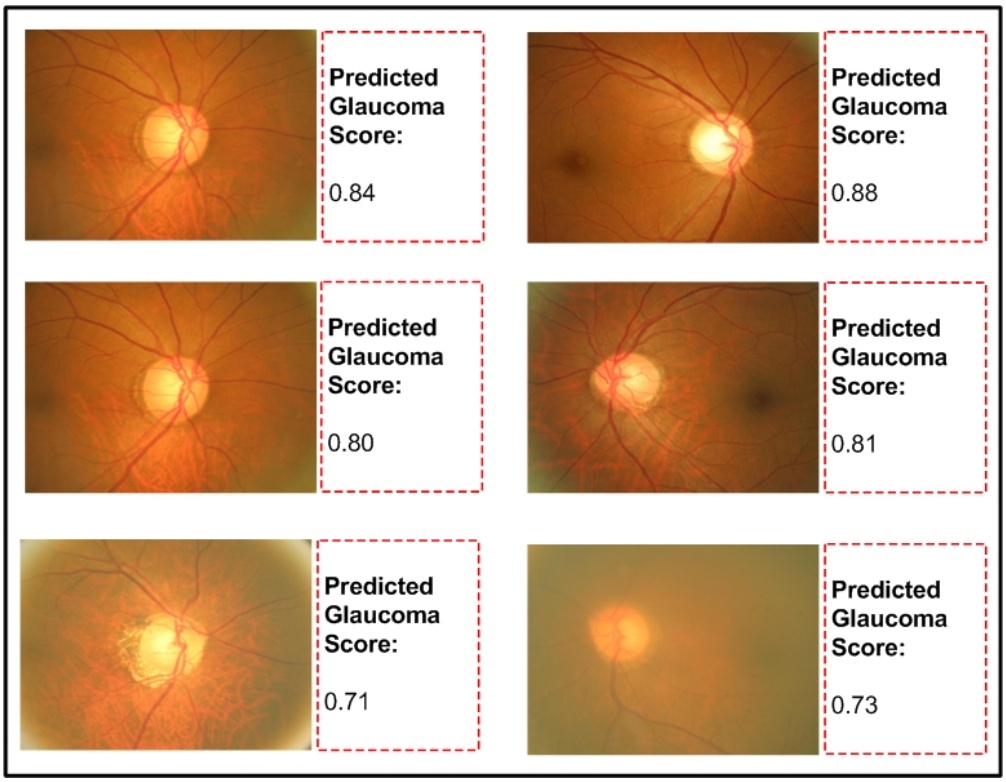

Fig. 2. Sample diagnosis results from our proposed algorithm for glaucoma detection. Each fundus image is diagnosed by clinicians and the predicted labels with probabilities by our algorithm.

V. Conclusion

In this paper, we present a DL framework for glaucoma detection based on deep CNN, which is able to capture the discriminative features that better characterize the hidden patterns related to glaucoma. The adopted DL structure contains six layers:four convolutional layers and two fully-connected layers. To reduce the overfitting problem, we adopt response-normalization layers and overlapping-pooling layers. In order to further boost the performance, dropout and data augmentation strategies are utilized in the proposed deep CNN. In future work, we plan to extend our study of DL architecture based on CNN to multiple ocular diseases detection.

References

[1] Zhang, Z., Yin, F., Liu, J., Wong, D.W.K., Tan, N.M., Lee, B.H., Cheng, J., Wong, T.Y.: Origa-Light: An Online Retinal Fundus Image Database for Glaucoma Analysis and Research. In: IEEE Int. Conf. Engin. in Med. and Biol. Soc., pp. 3065–3068 (2010)

[2] Wong, D.W.K., Lim, J.H., Tan, N.M., Zhang, Z., Lu, S., Li, H., Teo, M., Chan, K., Wong, T.Y.: Intelligent Fusion of Cup-to-Disc Ratio Determination Methods for Glaucoma Detection in ARGALI. In: Int. Conf. Engin. in Med. and Biol. Soc., pp. 5777–5780 (2009)

[3] Zhang, Z., Lee, B.H., Liu, J., Wong, D.W.K., TAN, N.M., Lim, J.H., Yin, F.S., Huang, W.M., Li, H.: Optic disc region of interest localization in fundus image for glaucoma detection in argali. In: Proc. of Int. Conf. on Industrial Electronics & Applications, pp. 1686–1689 (2010)

[4] Sng, C.C., Foo, L.L., Cheng, C.Y., Allen, J.C. Jr, He, M., Krishnaswamy, G., Nongpiur, M.E., Friedman, D.S., Wong, T.Y., Aung, T.: Determinants of Anterior Chamber Depth: the Singapore Chinese Eye Study. Opthalmology 119(6), 1143–1150 (2012)

[5] Michael, D., Hancox, O.D.: Optic disc size, an important consideration in the glaucoma evaluation. In: Clinical Eye and Vision Care 1999

[6] Harizman, N., Oliveira, C., Chiang, A., Tello, C., Marmor, M., Ritch, R., Liebmann, J.M.: The isnt rule and differentiation of normal from glaucomatous eyes. In: Arch Ophthalmol 2006

[7] Xu, Y., Lin, S., Wong, T.Y., Liu, J., Xu, D.: Efficient Reconstruction-Based Optic Cup Localization for Glaucoma Screening. In: MICCAI 2013

[8] Liu, J., Wong, D. W. K., Lim, J. H., Li, H., Tan, N. M., Zhang, Z., Wong, T. Y., Lavanya, R.: ARGALI: An automatic cup-to-disc ratio measurement system for glaucoma analysis using level-set image processing. In: 13th International Conference on Biomedical Engineering (ICBME) 2008

[9] Le, Q.V., et al.: Building high-level features using large scale unsupervised learning. In: ICML 2011

[10] Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. In: Arxiv 2012

[11] Ciresan, D.C., Meier, U., Masci, J., Gambardella, L.M., Schmidhuber, J.: High-performance neural networks for visual object classification. In: Arxiv 2011

[12] Xu, Y., Xu, D., Lin, S., Liu, J., Cheng, J., Cheung, C.Y., Aung, T., Wong, T.Y.: Sliding Window and Regression based Cup Detection in Digital Fundus Images for Glaucoma Diagnosis. In: MICCAI 2011

[13] Xu, Y., Liu, J., Lin, S., Xu, D., Cheung, C.Y., Aung, T., Wong, T.Y., Efficient Optic Cup Detection from Intra-image Learning with Retinal Structure Priors. in : MICCAI 2012

[14] Jonas, J.B., Fernandez, M.C., Naumann, G.O.: Glaucomatous parapapillary atrophy occurrence and correlations. In: Arch Ophthalmol 1992

[15] Bengio, Y., et al.: Representation learning: A review and new perspectives. In: Arxiv 2012

[16] Quigley, H.A., Broman, A.T.: The number of people with glaucoma worldwide in 2010 and 2020. In: Ophthalmol 2006

[17] Damms, T., Dannheim, F.: Sensitivity and specificity of optic disc parameters in chronic glaucoma. In: Invest. Ophth. Vis. Sci. 1993

[18] Krizhevsky, A., et al.: Imagenet classification with deep convolutional neural networks. In: NIPS 2012